Homework - Day 7

Author: Daniel Ramírez, 2022

Updated by Samuel Hunter, 2024

DESeq2 resources:

https://bioconductor.org/packages/release/bioc/html/DESeq2.html
https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html

Introduction:

Part 1: FeatureCounts
Practice generating counts using the example bam files on the AWS

1) Find the FASTQ files in `/scratch/Shares/public/sread2024/homework_data_files/day7/fastq/`
    a. Process these FASTQ files into BAM files
    b. Using the hg38 GTF file in /scratch/Shares/public/sread2024/data_files/project/day7/annotations/hg38_ucsc_genes_chr21.gtf, count the reads for all genes in the GTF file
2) Check the output. How many samples did you get reads for? How many genes?
    a. Check the gene ***UBE2G2***. How many reads did each sample get for this gene?

Part 2: DESeq2

Andrysik et al. ran several experiments which identified a core regulatory program associated with p53 activation across multiple cell lines. In class, we ran differential analysis on RNA-seq data in HCT116 cells. Here, you will run that same pipeline on another cell line from the same paper.

`srworkshop/projectB/day07/homework/featureCounts/MCF7_counts.tsv`
`srworkshop/projectB/day07/homework/featureCounts/MCF7_samples.tsv`

1) Read these files into your R environment. Are these files in the proper format to enter into DESeq2?
2) Run DESeq2 on these samples, using an experimental design that tests whether the Nutlin-treated samples show any significant differences from the DMSO-treated ones
    a. Generate histograms and boxplots for the normalized counts.
    b. Generate a PCA plot, coloring the samples by their treatment group. How do the samples group together?
3) Generate the DESeq2 statistical results. Use an adjusted p-value cutoff of 0.1.

a. How many genes were upregulated upon Nutlin treatment? How many were downregulated?
b. Generate an MA plot and a Volcano plot. Color the significant genes.
c. What's the top hit in the DESeq2 results?